

# A Review of the ISLSCP Initiative I CD-ROM Collection Context, Scope, and Main Outcome

*by Yann H. Kerr, CESBIO/LERTS*

*with contributions from*

*Peter Briggs, Jim Collatz, Gerard Dedieu, Han Dolman, John Gash, Forrest Hall,  
Alfredo Huete, Fred Huemmrich, John Janoviak, Randy Koster, Sietse Los,  
James McManus, Blanche Meeson, Ken Mitchell, Michael Raupach, Piers Sellers,  
Paul Try, Ivan Wright, and YongKang Xue.*

---

## CONTENTS

### **I. Overview**

### **II. General Outline of the Review Process**

- 2.1 Stage One: Documentation Review
- 2.2 Stage Two: Qualitative Analysis of the CDs
- 2.3 Stage Three: “Hardware” Review of the CDs
- 2.4 Stage Four: Extensive and Quantitative Review of the CDs

### **III. Qualitative Review of the Initiative I CD Collection**

- 3.1 Scope of the Review
- 3.2 Charge to the Reviewer
- 3.3 Organization of the Review
  - 3.3.1 *Data Types*
  - 3.3.2 *Methodology*

### **IV. Output of the Review**

- 4.1 Vegetation: Land Cover and Biophysics
- 4.2 Hydrology and Soils
  - 4.2.1 *Precipitation*
  - 4.2.2 *Soils*
  - 4.2.3 *Runoff*
- 4.3 Snow, Ice, and Oceans
- 4.4 Radiation and Clouds
  - 4.4.1 *Radiation*
  - 4.4.2 *Albedo*
  - 4.4.3 *Clouds*
- 4.5 Near-Surface Meteorology

### **V. Conclusion**

# I. Overview

A CD collection of global data sets has been issued within the framework of ISLSCP Initiative I. The rationale for producing this CD set is described in P. J. Sellers, et al. (Remote sensing of the land surface for studies of global change: Algorithms, model, experiments. *Rem. Sens. Environ.* 51:1:3-26). This collection should be of considerable interest to land-atmosphere modelers since data sets such as these are difficult to obtain in one package. However, there are risks involved in releasing such a collection. Scientists may consider its contents “gospel” (especially when the data come from another scientific community) and may misuse the data or reject them as worthless or grossly wrong, which could discredit the ISLSCP Initiative in its entirety. Consequently, releasing the collection with insufficient explanation had to be avoided.

Thus, the ISLSCP Science Steering Committee decided to review the different data sets and include the results of the review (this text) on the CDs. Because of time constraints, only a qualitative analysis, not a full review process, was performed. In most cases the review consisted of looking at a subsample of the different data sets (1 or 2 months, generally January and July), identifying obvious problems, and suggesting corrections. Most of the corrections were made but not reviewed. Real intercomparison of similar data sets was not performed.

Generally speaking, the review showed that the data had the correct “look and feel.” All reviewers agreed that, despite some problems, these CDs were very useful and almost always superior or equal to existing, though scattered and often inaccessible, data sets.

As you will most probably use one or several data sets included in this collection, you may come up with relevant comments and a more quantitative analysis of the contents. Consequently, we welcome all your comments toward producing a more quantitative statement of worthiness and an improved Initiative II data set collection on CDs. These should be sent to the editors of the CD collection (Blanche Meeson, Code 902.2, NASA Goddard Space Flight Center, Greenbelt, MD 20771. Email: meeson@eosdata.gsfc.nasa.gov. Voice: 301-286-9282).

## II. General Outline of the Review Process

Time constraints necessitated splitting the review process into four stages as follows:

### 2.1 Stage One: Documentation Review

During the September 1993 ISLSCP Science Steering Committee (SSC) meeting, it was decided to have the documentation reviewed separately to identify errors and omissions of material essential for novice users. Tasks also included flagging ranges of validity, main sources of errors, relevant literature, and integrity of the documentation; reviewing for clarity, completeness, and data comprehension; and checking the data format description and data acquisition information. This review led to improved documentation files—this is actually the main core of the CD review for the ground data sets and the data sets with a long track record (e.g., Surface Radiation Budget (SRB) data sets).

### 2.2 Stage Two: Qualitative Analysis of the CDs

The role of this stage is detailed in Section III. The output is given in Section IV.

### 2.3 Stage Three: “Hardware” Review of the CDs

This review consisted of a quick look at a test version of the CD set (issued in limited number: under 15 copies). The reviewers were expected to check that the CDs are readable, the data are organized correctly, and everything is present and similar to the original data set they reviewed as separate items sent to them via e-mail or FTP files, etc. This was done.

### 2.4 Stage Four: Extensive and Quantitative Review of the CDs

This stage is for you, the user, to do. Please return your comments and opinions to the editors of this CD collection (Blanche Meeson, Code 902.2, NASA Goddard Space Flight Center, Greenbelt, MD 20771. E-mail: meeson@eosdata.gsfc.nasa.gov. Voice: 301–286–9282). We ask that you

- a) include relevant information you gathered while using the CDs by comparing the data sets on the CDs with related

data sets such as your own, model output, and large-scale experiment results

- b) suggest improvements, flag doubtful data, analyze the processing steps, etc.—the data sets were intended to cover all of Earth's biomes; we are sure that the quality of some of the products will vary with geographical location and perhaps season.

This final review could culminate in a publication in the open literature and maybe a workshop within a year or two of release of this CD collection. The ISLSCP Science Steering Committee would also analyze the outcome of this second review in light of Initiative II products and their ongoing program of reviews.

## III. Qualitative Review of the Initiative I CD Collection

### 3.1 Scope of the Review

The Initiative I goal was to produce CDs containing “available” state-of-the-art data sets and products from reliable and readily available data. Toward this purpose, reviewers were asked to check the validity and usefulness of the data sets; identify caveats, doubtful parameters, and big mistakes; and assess validity ranges, glaring gaps, and redundancies with other data sets. When applicable, they also suggested improvements.

### 3.2 Charge to the Reviewer

*Perform a qualitative analysis (quantitative analysis whenever possible) of the data sets from your knowledge of the discipline, cross comparisons with other similar data sets, model output, etc.*

For this purpose, the reviewer was expected to check the data sets falling into his area of expertise and personal knowledge and answer the following questions.

- Which is the area or biome where I have made my comparison? Are the data grossly wrong or do they compare well with what I have seen or measured?
- From my experience, how accurate (error bars) are the data?
- If differences are found, what are the possible explanations?
- What is the validity range of the data (i.e., range of physical values, geographical areas, perturbing factors)?
- What are the caveats or limitations of the data?
- Are all the parameters relevant or useful?
- What other similar data sets exist?
- What are the temporal and spatial sampling characteristics? Do they accurately reflect reality (representativity)? Do they affect the usefulness of the data?
- For “processed” data, what is my opinion about the processing steps, assumptions made, and impact on the output quality?
- Do I have any suggestions for improvements?

### 3.3 Organization of the Review

The first step is to distinguish between the different types of data sets, since the review process might differ from one type to another.

#### 3.3.1 Data Types

We identified three types of data sets: satellite data, ground data, and model output. Merged data sets were considered in both categories; for example, if a data set contained ground and satellite data it was reviewed as both categories.

*a) Satellite data—two subcategories:*

- those that are new to the research community—the review process concentrated on two topics:
  - 1) analysis of the methodology used to process the data (identify caveats, oversimplistic or wrong assumptions, etc.)
  - 2) comparison of these products to other data sets (ground, model, experiments)—the data sets in this subcategory were the most important to review.
- those having a long track record, such as the Surface Radiation Budget data sets—these were reviewed similarly to model output data sets.

*b) Ground data*—For the most part, ground data had to be taken as given. The review focused on the known limitations of measurement techniques, sampling (temporal and spatial), representativity, and accuracy. Where ground data were produced after some processing steps, the reviewers were asked to give their opinion about the procedures used.

*c) Model output*—The main scope of the review was to identify questionable model products, range or area of validity, usefulness or relevance of the different parameters, comparison with ground data and satellite data, accuracy and reliability, main limitations, and known problems. The problem with model output products is that they are usually self-consistent and have the look and feel of actual data. Novice users tend to consider these data as “truth,” whereby specialists are aware of the limitations.

#### 3.3.2 Methodology

The first step (Stage One) was to send the documentation to the document reviewers for a thorough review of content and accuracy. When the document

reviewers' comments were received and incorporated, the data and documentation were sent to the Stage Two reviewers.

Stage Two reviewers were sent the documentation along with reviewing instructions via electronic mail. These documents introduced the data sets that were to be reviewed and delineated the scope, charges, and schedule of the review. The reviewers were then sent the data sets via FTP from Goddard Space Flight Center (GSFC).

When the data had been reviewed, a meeting was held at GSFC (October 26-27, 1994) during which the data sets were analyzed by qualitative analysis of a couple of samples from each data set. The outcome of this review was a whole set of suggested improvements and harmonization of notations. In several cases, alternative data sets were suggested, and, for lack of availability or too poor quality, some data sets were replaced with others. A second meeting took place January 4, 1995, at GSFC. During this meeting the added data sets and the corrections made to the first-round data sets were checked. The final output of the review process is given in Section IV.

Once all the data sets were completed, test CDs were produced and checked to ensure that the data were properly encoded on the CDs (Stage Three, March 1995).

## IV. Output of the Review

### 4.1 Vegetation: Land Cover and Biophysics

The satellite data were divided into two categories: data sets having a long track record (see 4.4) and “new products.” Both types show caveats, but it was considered that they needed pointing out only in the latter case. The vegetation land cover and biophysics data set was analyzed as a satellite data set in the category “new to the community” (cf. 3.3.1). It is the most challenging data set to review and one of the most interesting on the CD set thanks to the global coverage of several parameters of interest for the modeling community. The user should well be aware, however, of the limitations of this suite of parameters. The limitations are linked mainly to the following facts.

- a) Nearly the whole data set is obtained from Normalized Difference Vegetation Index (NDVI) data. The input data consist of NDVI (2 years), a vegetation map derived from NDVI data, and Earth Radiation Budget Experiment (ERBE) data over the lower latitudes. Consequently, we have only two really independent data sets in some areas and one in others, with added specific information (respiration, C3/C4 etc.; see VEG\_CLSS.DOC in the Documents folder on the CD).
- b) Any mistake or error in, say, the vegetation map will consequently propagate in all related output files: check closely the validity over your area of interest (Scotland seems to be covered with forest, for instance).
- c) NDVI is used with all the limitations of this quantity. No atmospheric corrections were done, but there were plenty of empirical procedures used to suppress problems linked with cloud cover.

This leads to constant values over rainforest, for example, throughout the year (one value is retained as good per pixel and kept for the whole year). Consequently, there is a “jump” (not necessarily significant, though) on December 31.

The Fourier transform tends to smooth the curve and suppress anomalies during vegetation growth (a decrease during the growing season due to a drought for instance) or smooth out or suppress short term evolution (i.e., semiarid fallow).



Sun angle correction is performed in a crude way (no relevant information available). It might cause problems around the equinoxes and along the scan.

In one direct comparison of these data with a higher resolution data set gathered over the FIFE site, these NDVI product values appeared to have lower than expected values in the middle of the growing season.

- d) The relationship to extract Fraction of Photosynthetically Active Radiation (FPAR) from Simple Ratio (SR) has been established over the Konza prairie. But, for many biomes, shadowing effects lead to a much less linear curve and, consequently, the obtained FPAR is sometimes largely underestimated.
- e) For defining background reflectances, ERBE data are pasted in areas of sparse or no vegetation and the limit is visible in some places (central Europe) as the values differ (largely in this case) from those obtained by assigning values by vegetation type as defined from analyses of NDVI.

Thus, for Initiative II it is strongly recommended that a more suitable input data set is used (actual reflectances and information on viewing and solar angles so that artificial cleanup methods are reduced to a minimum). Basic atmospheric corrections could then be done with water vapor from the European Center for Medium-Range Weather Forecast (ECMWF) or similar data. SR-FPAR relationships should be more thoroughly tested. It was also suggested to use Bidirectional Reflectance Distribution Function (BRDF) models.

## 4.2 Hydrology and Soils

The data sets in this category were analyzed as “ground data” and “merged data sets.” Generally speaking, ground data have been the most difficult to gather. After the review process it was decided to drop several data sets originally considered for inclusion in this CD collection because they were too unreliable or the coverage of land surfaces was too sparse to be of any use for global modeling. Those global ground data sets that appear on the CDs were always judged as useful or very useful, in spite of a sometimes questionable accuracy. They are the only source of global, uniform data accessible without the usual hassle.

#### ***4.2.1 Precipitation***

The monthly precipitation data set consists of data derived from analyses of surface gauge observations. The rainfall data set is the state of the art but might vary in quality with geographical location. This is due mainly to the spatial coverage available (some areas have a very sparse gauge coverage), and to the more basic problem of temporal sampling and representativity of values derived over a  $1^\circ \times 1^\circ$  grid from few, not regularly spaced ground measurements.

The representativity is fair temporally but slightly poor spatially, as one would expect. When compared with other (field campaign) measurements in Sahel and Brazil some discrepancies were found that were sometimes important. This is probably due to spatial representativity. Users should be aware of these possible variations and should check, over a given area of interest, whether the number of stations used over the  $1^\circ \times 1^\circ$  area is sufficient to give credible results. The CD collection also holds a merged monthly satellite-surface precipitation product at  $2.5^\circ \times 2.5^\circ$  resolution: this is continuous over the land and oceans and is provided only as a browse file.

A hybrid precipitation product was generated by using the NMC GCM analysis output and data from a large-scale observational program (GARP) to divide up the GPCP  $1^\circ$  monthly data set described above into 6-hourly total and convective precipitation amounts, which can then be used in conjunction with the ECMWF 6-hourly products. The accuracy of this hybrid product is unknown.

#### ***4.2.2 Soils***

The soil data set was put together from a variety of existing sources. It contains some information on soil composition, texture, depth, and slopes. It must be noted that these data sets are to be considered as is. They are not as accurate as desired, and the information content might not satisfy all users. It is, however, the state of the art, and it is thought to be not possible to get better information on a global basis at this time. To quote a reviewer, "There is no new information on the CDs, just a concatenation of existing data sets. So it is clearly a case of rubbish in, rubbish out." The advantage is that on a CD the presentation stops at the right point, that is, at the leaping off place where the qualified expert would not dare to go. The data on the CD are considered as better globally than other existing data sets. However, locally (Amazon Basin, in this case) it is only equivalent to existing data sets because of the poor source of data. Over the Amazon Basin, it was found that the texture data are fairly accurate, but when they were used to infer available soil mois-

ture they proved to be questionable. This probably applies to all areas of specific soils not well parameterized.

It must be noted that the slopes seem too high and that there is apparently a problem over Greenland where the slopes are greater than over the Great Cascade in Alaska. This is probably due to the way the slopes are computed from a data set containing only three ranges. For Initiative II, the slopes will probably have to be directly estimated from a digital elevation model. For similar reasons, the soil type data set has limitations linked to the input data.

#### ***4.2.3 Runoff***

The runoff data set also suffers from several gaps. From a total of 34 basins, only 14 are available for both 1978 and 1988. The consistency of the flow rates is not very good. This data set should be used for checks since the coverage is not global and not fully reliable. It should not be used as input data. The efforts for Initiative II will probably have to concentrate on improving these ground data sets.

### **4.3 Snow, Ice, and Oceans**

These data sets are to be taken as is and considered with much care. One should first notice that the NOAA/NESDIS data set (snow extent) covers only the northern hemisphere. Some doubtful results were also found (USAF ETAC snow depth) over Greenland (very high snow depth) and the Snow Cover Data Set has some isolated anomalies; e.g., New Zealand (snow in January!). Some problems were found also while regridding the polar stereo projection to the standard grid used on the CD.

The ocean data sets were not reviewed.

### **4.4 Radiation and Clouds**

These data sets were analyzed as “satellite data with a long track record.” Users should refer to the documentation file for possible caveats, terminator effects, and so on.

#### ***4.4.1 Radiation***

There are several data sets of interest in this category: the ECMWF data (see 4.5), ERBE data, Staylor and Darnell (Langley Research Center), and Pink-

er's. The main problem encountered was satellite coverage that did not cover the complete globe. Gaps were filled through an interpolation method (Pinker) after the first review. Nevertheless, discrepancies occur at the limits of the coverage of the different geostationary satellites. It was also found that the radiation values were compatible with climatological values with differences of between 10 and 20 percent in some cases, which can be attributed to sampling and interpolation problems in the climatological data sets. In the Sahel area, the ISLSCP radiation values well captured the seasonal variability ( $\pm 20^\circ$  W), while in the Amazon Basin, only longwave down and net longwave agreed with ground measurements. The shortwave down appeared bad, and the shortwave net and total net were not very good. Moreover, the seasonality of the signal found on the ISLSCP data set ( $\pm 70^\circ$  W) is not visible on ground measurements. A registration error in the second part of 1987 was detected. Globally the radiation data seem reasonably accurate with some local problems that are largely compensated by the available global coverage. For Initiative II it was recommended to improve the aggregation technique. In addition, it was recommended that the authors be less vague on the description of their procedures in the documentation file.

#### ***4.4.2 Albedo***

There seem to be some registration errors in the ERBE data set ( $5^\circ$  W), at least for some months. Significant differences were also found between the ERBE Top of Atmosphere (TOA) albedos and the Langley surface values (higher over the oceans and lower over the land), but it was also found that the ERBE albedo was slightly too high (over Sahel and Amazon). It is recommended that the documentation file clearly describe the differences between TOA and surface so that the uneducated user has some views on that problem.

#### ***4.4.3 Clouds (and Atmospheric Data)***


This data set (International Satellite Cloud and Climatology Project) was put on CD as is. It has several problems due mainly to the different algorithms used over sea and land (the continent contours are visible!) and to the imperfect intercalibration of the different sensors or gap filling procedure (vertical structure west of the Indian subcontinent linked to METEOSAT and GMS coverage). Finally the values at the extreme latitudes seem erroneous (cloud water for instance) and Greenland looks strange. This data set has to be used with much care. Some reviewers suggested discarding the cloud optical thickness and cloud water path, but it is included here because others thought it essential.

## 4.5 Near-Surface Meteorology

The output of the review is very small on this data set. The problems were

- a) It is very difficult to check and there was really only one data set available at the beginning of this CD initiative (ECMWF). Model outputs are of the self-consistent type. The model runs with various assumptions (sometimes gross) so that the output of directly useful or checkable products makes some sense. Consequently, some output data do not make much sense. The ISLSCP SSC and review team did some “pruning” of seemingly worthless data and decided to elaborate new products of use in modeling (see documentation files) from existing data.
- b) The data set arrived late and proved to be difficult to process, and contained a large volume of data (four out of the five CDs). Thus, we did not have much opportunity to go through it. Consequently, the data sets are to be considered as state of the art to be taken as is, but not necessarily as gospel. The user is strongly encouraged to read the documentation file carefully and if not a “trained user” to ask a modeler in case of doubt.

## V. Conclusion



The CD collection review process has been an interesting and valuable experience. We believe that it has enabled a significant improvement of the content. Our only regret is that the time constraint has been too strong, not allowing the reviewers to go as deep as they would have liked in the analysis. We believe, nevertheless, that users will provide us with their comments so that a more complete review will eventually emerge, and the Initiative II CD collection will benefit from user feedback and a more indepth review.

Finally, the reviewers are deeply indebted to Blanche Meeson and James McManus who, with very short notice, made this review possible in spite of various and complex problems they had to solve to put together these data sets and the reviewers' "suggested" changes.